

基于主题图的非物质文化遗产数字资源整合实证分析*

■ 施旖¹ 熊回香² 陆颖颖²

¹ 武汉大学图书馆 武汉 430072 ² 华中师范大学信息管理学院 武汉 430079

摘要: [目的/意义] 非物质文化遗产资源具有重要的社会价值,将主题图技术应用于非物质文化遗产数字资源的整合中,能够更好地实现非物质文化遗产数字资源的传承与保护。[方法/过程] 在分析非物质文化遗产数字资源的主题、关联关系和资源标引的基础上,构建非物质文化遗产数字资源主题图并实现其可视化。[结果/结论] 研究结果显示,主题图技术的应用可以为非物质文化遗产数字资源提供基于语义的整合,并将主题之间的关联直观地展示给用户,为用户提供可视化导航。

关键词: 主题图 非物质文化遗产 数字资源 可视化

分类号: G254

DOI: 10.13266/j.issn.0252-3116.2018.07.012

1 引言

非物质文化遗产(以下简称“非遗”)和物质文化遗产都是人类社会文明的伟大结晶,具有重大的社会价值。和物质文化遗产的可见性对比,非遗都是由口、肢体、行为和技能来表达的,具有无形性和活态性^[1],非遗从产生到发展并不孤立存在于人的精神世界,而是和它所处的生态空间和文化环境紧密相连,只有在它顺应的环境中才具有活态性^[2]。这让非遗的保护和传承面临更多困难。目前,国内外学者开始利用元数据、本体、主题图、关联数据等知识组织技术去探索非遗的语义化组织,如非遗元数据标准的建立和比较、非遗本体的构建、非遗主题图的构建,以及基于关联数据的非遗资源聚合平台的搭建等。然而,在众多非遗资源组织的方法中,关于主题图应用于非遗的研究相对较少,通过在知网、万方、Web of Science 等数据库中以“非物质文化遗产”“非遗”“主题图”为关键词进行检索,发现仅有两篇文章研究了主题图与非遗的结合^[3-4]。此外,有关非遗资源的语义组织虽有一定研究,但仍存在一些问题,例如,元数据侧重于对主题概念进行描述,难以描述主题间的关联关系;本体侧重描述概念之间的关系,不能描述主题与资源实体间的关联,且构建较为复杂;关联数据侧重描述主题与资源实

体之间的关联,但使用前提是将所有资源转化成统一的数据格式。主题图作为一种新的数字化知识组织技术,吸收了语义网和本体的思想,可以创建丰富的语义模型,浏览和检索各种信息,整合各种不同的数据资源,具有很强的灵活性,有知识导航、知识定位等多种功能。因此,笔者将主题图技术应用到非遗数字资源组织研究中,构建非遗数字资源主题图,并以发源于湖北和安徽地区的“黄梅戏”为例进行实证研究。研究发现,主题图应用于非遗数字资源的整合,不仅可以更充分地实现非遗的利用价值,有利于其保护和传承,还能以可视化的方式直观地展现非遗的历史源头、传承脉络、流派发展等信息,为用户的利用提供了更加合理、有效的途径。

2 主题图及非遗数字资源整合概述

2.1 主题图技术

ISO/IEC 13250 对主题图进行了以下介绍:主题图定义了一个多维空间,这个空间中分布着许多结点,每个结点代表一个主题,主题间的距离可以通过从主题到主题所必须经过的中间主题的数量来计算,并且这些中间主题构成了这两个主题之间建立关系的路径^[5]。主题图有三个基本要素,分别是主题、关联、资源指引^[6]。其中,“主题”可以是任何事物,无论其存

* 本文系国家社会科学基金项目“大众分类中标签间语义关系挖掘研究”(项目编号:12BTQ038)研究成果之一。

作者简介:施旖(ORCID:0000-0002-8952-6126),助理馆员,硕士;熊回香(ORCID:0000-0001-9956-3396),教授,博士,通讯作者,E-mail:hxxiong@mail.ccnu.edu.cn;陆颖颖(ORCID:0000-0001-6556-0621),博士研究生。

收稿日期:2017-11-07 修回日期:2018-01-21 本文起止页码:104-110 本文责任编辑:王善军

在与否都可以是主题,一个主题可以有一个或多个基本名称,还可以有别称。“关联”代表两个或多个主题之间的关系,可以被划分为不同的类型,即关联类型。一个主题可以连接与它相关的一种或多种信息资源,这些与其相连接的信息资源就叫做资源指引,也有学者称其为“资源实体”“资源出处”等,主题图内的资源叫做内部资源指引,由统一资源定位符 URI 链接的资源称作外部资源指引。

主题图的概念模型包括两层,分别是主题地图层和信息资源层。主题地图层由各主题之间的关联构成,如图 1 所示。图 1 中的一个圆圈代表主题,其间的连线表示主题间的关联。信息资源层包含各种信息资源,这里的信息资源通常指数字信息资源,如一篇文章、一个网页、一个视频等等。这些资源与主题地图层中的主题相联系,又独立地存在于互联网。

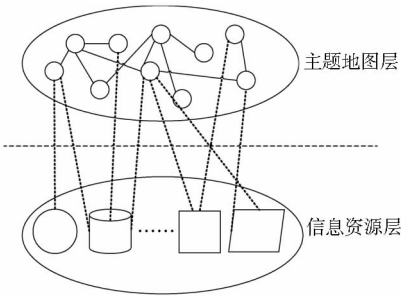


图 1 主题图概念模型

2.2 非遗数字资源整合

联合国教科文组织在《非物质文化遗产公约》定义非物质文化遗产为个人、集体或社区视为文化遗产的各类知识、表演、技能、社会实践、观念表述、表现形式及相关手工艺品、实物、工具和文化场所。非遗数字资源指结合计算机技术、通信技术、多媒体技术而形成的以数字化形式发布、存取和利用的非遗资源,如非遗音频资源、非遗视频资源、非遗数据库等。非遗数字资源整合是根据非遗的分类体系,通过对非遗数字资源的内部特征和外部特征进行表征与排序,从而实现非遗数字资源从无序向有序的转变,促使非遗数字资源形成科学有序的整合。非遗数字资源不是一个独立的资源个体,它是一个地方独特的文化资源特色,必须借助文化传承人将其发扬光大,所以说非遗与其所处的生态空间和文化环境紧密相连。为了完整地展示非遗数字资源及资源间的关系,需要对非遗数字资源进行语义组织,即按照非遗的分类体系,对非遗数字资源进行主题的抽取,确定主题之间的关系,并对各类非遗数字资源之间的关系进行规范描述,形成一个基于语义

的非遗数字资源整合模型。

对主题图和非遗数字资源整合相关理论的梳理,有助于更好地把握各自的特点和两者之间的关系,从而在非遗数字资源的整合中充分利用主题图的优势,有利于基于主题图的非遗数字资源整合模型的构建。

3 基于主题图的非遗数字资源整合模型构建

主题图最大的优势主要体现在它将资源实体与主题相分离,非遗数字资源主题图的构建是在采集非遗数字资源的基础上抽取相关主题概念,然后建立主题概念间的关系,最后将这些主题概念和非遗资源实体进行关联。笔者依据主题图构建的基本原理和非遗数字资源的相关知识构建了一个基于主题图的非遗数字资源整合框架,并分析了主题图应用于非遗数字资源整合的优势。

3.1 基于主题图的非遗数字资源整合框架

依据主题图的基本原理,可以构建非遗数字资源整合框架,如图 2 所示。它主要包括 4 个部分,分别是资源层、生成层、应用层和用户层^[7]。

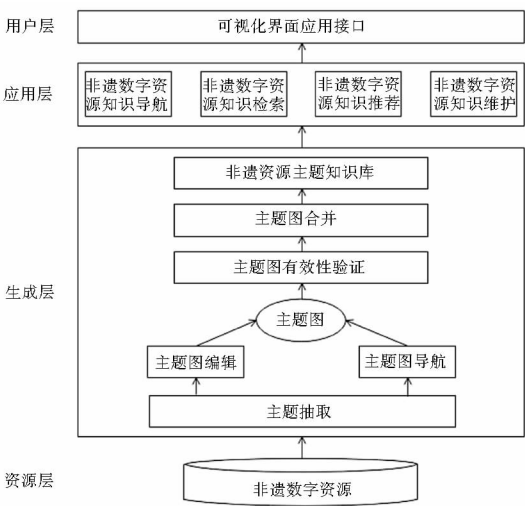


图 2 基于主题图的非物质文化遗产数字资源整合框架

3.1.1 资源层 非遗相关资源是完整展示非遗数字资源语义结构的资源基础,资源层在基于主题图的非遗数字资源整合框架的最底层,是整个资源整合的起点,它为主题图的构建提供大量非遗相关的异构信息资源,包括数字资源、与非遗相关的人员资源、机构资源、知识资源、项目资源等等。

3.1.2 生成层 主题图生成层是将搜集和整理的非遗数字资源进行主题概念的抽取,划分主题类型、关联

和资源指引,再使用 Ontopoly 组件编辑主题图的本体和实例,对非遗数字资源主题图本体进行语义标注,接着检验主题图表示非遗数字资源的正确性,分析其中的关联关系,合并具有同样关联的主题图,最后将有效的主题图存储到非遗数字资源知识库中。

(1)定义主题类型。非遗数字资源涉及范围广泛,种类繁多,因此需要对其进行主题划分。笔者从层级角度出发,将非遗数字资源分为 6 类一级主题,分别为“民族”“地域”“非遗项目”“级别”“传承人”“申报者”,之后再根据各主题实际情况,依次划分各子主题。其中,“民族”表示非遗项目的民族特色,“地域”表明了地域划分,“非遗项目”确定了非遗的项目名称,“级别”表现了非遗的重要程度,“传承人”是非遗的传承历史及主要传承人物,“申报者”是非遗项目的申报主体。非遗数字资源包含的种类众多,内容也十分丰富,且不同领域之间的差别较大,而构建一个完整的非遗数字资源主题图需要投入大量的精力和成本,因此笔者选取湖北省申报的国家非遗项目“黄梅戏”,构建黄梅戏数字资源主题图。黄梅戏为非遗项目的其中一种,在一级主题的基础上,再针对黄梅戏自身特点,分别设置其二级主题、三级主题等。

(2)定义关联类型。定义了主题类型之后,就要对主题间的关联进行定义,主题间的关联将不同的主题联系起来,形成一个知识网络。笔者针对前面定义的 6 种主题类型定义了 6 个关联类型,分别为“民族和非遗”“非遗项目、申报者、级别”“师徒”“传承”“位于”“属于”关联。其中,“民族和非遗”是民族与非遗项目之间的关联关系,“非遗项目、申报者、级别”是非遗项目、申报者与级别之间多对多的复杂关系,“师徒”是传承人之间的传承关系,“传承”是传承人与非遗项目之间的传承关系,“位于”是非遗项目与地域之间的关系,“属于”是一级主题与子主题之间的关系。二级主题的关联主要包括“包含/属于”关系、“近义”关系、“合作”关联、“演绎”关联、“师徒”关联、“创作”关联。

(3)资源指引。资源指引主要是对主题的资源指引与定位,根据资源内容的不同,在非遗相关资源的基础上,把资源类型划分为非遗数字资源、机构资源、人员资源、项目资源 4 个类别。其中,数字资源按其资源表现形式分为文本、图片、音频、视频、模型 5 类。

3.1.3 应用层 应用层是在知识服务过程中,根据用户需求对非遗数字资源知识内容进行检索、重组、查询,从而为用户提供所需的服务。应用层主要提供检

索、导航、推荐、维护等功能。检索功能,用户可以通过主题图查询语言查找感兴趣的主体,定位所需的主题及其相应的资源指引;导航功能,主题图中主题之间的关联关系可以为用户提供导航帮助,用户可以借助导航进行拓展检索;推荐功能,当用户检索某一主题时,主题图的语义关联特性将为用户自动推荐与用户检索主题有关联关系的相关资源,提高用户检索资源的查全率;维护功能,非遗主题图的知识维护功能包括增加、删除、更新三种基本操作方法和主题、主题类型、关联关系三种基本要素,通过该模块可以对非遗主题图进行不断地完善和维护。

3.1.4 用户层 用户层处于非遗数字资源主题图体系结构的最顶层,为用户提供主题图浏览、检索等功能,用户可利用主题图的 Omnigator 组件浏览主题知识页面,并使用用户接口调节各个模块。

3.2 基于主题图的非遗数字资源整合的优势分析

主题图被称作信息时代的 GPS^[8],是一项可以实现知识管理的技术,在非遗数字资源的整合中有其独特的优势。就非遗数字资源整合方面,主题图的优势体现在如下几个方面:

第一,语义描述方面。主题图起源于传统的查询辅助工具,如书后索引、叙词表等,它汲取了传统索引、图书馆科学知识表示的优点,可以被用来描述复杂的知识结构^[9]。主题图中的主题和关联允许灵活地定义复杂的语义关系,为非遗数字资源的组织提供一个基于语义的接口,用户可以实现基于语义的非遗数字资源查询检索。

第二,整合分布式资源方面。主题图利用资源指引达成非遗数字资源与主题间的链接,通过这种映射关系可以实现对广泛分布的异构非遗数字资源的索引,能有效解决目前非遗数字资源的信息孤岛现象。

第三,知识组织方面。主题图基于主题对知识分类,通过主题之间的关系将非遗数字资源的知识体系结构化,使得我们可以从知识层面来组织非遗数字资源,并建立非遗数字资源知识库。

第四,结果可视化方面。主题图提供了可视化的效果呈现方式,图像的方式使用户对非遗数字资源的主题及其关联关系一目了然。其中,主题图既可以选择呈现整体效果,让用户对非遗数字资源有总体的了解,也可以选择调整局部控制参数,观看局部效果。同时,当用户检索指定关键词时,检索出的关键词颜色会发生变化,方便用户点击。

联。

“包含/属于”关联。“包含/属于”关联就是黄梅戏主题之间存在的上下位关系,主题类型和其包含的子主题类型或子主题之间就存在这种关联关系,一些不同层级的主题之间也存在这种隶属关系。本文中存在的“包含/属于”关系的有:“剧目”主题类型包含“唱段”,不同的“剧目”包含各自的“脚色行当”,“唱段”属于不同的“音乐风格”。

“近义”关联。黄梅戏的有些主题之间存在着近义关系,比如黄梅戏音乐风格中“悲恸”和“悲切”、“抑郁”和“忧郁”、“欢快”和“活泼”之间即存在近义关系。

“合作”关联。出演同一个剧目的不同表演者,同一场表演中的导演、编剧、演员之间都存在合作关系。

“演绎”关联。表演者和他出演的作品、出演的角色行当之间存在演绎关系,因此可以在“表演艺术家”“剧目”“电影”“电视剧”“舞台剧”“脚色行当”之间构建演绎关联。

“师徒”关联。黄梅戏的传承方式是传统的拜师学艺方式,主要是以“以师带徒,师徒传承”的方式传承的。因此,许多黄梅戏艺术家之间都具有“师徒”关系,如严云高是严凤英唯一办过师仪的师傅,严凤英自己也收了田玉莲、秦启霞、王凤枝等人为徒。

“创作”关联。导演、编剧、曲作家等和他们的论文、文章、图书等文献、影视、音乐等之间存在创作关系。

4.3 资源指引分析

本文黄梅戏主题的相关资源链接主要来源于中国非物质文化遗产网、中国曲艺网、中国知网、万方数据库、非物质文化遗产名录数据库和搜索引擎等。在描述主题时,对主题属性加以介绍并标引相关资源可以让用户更加了解主题的意义和内涵,主题图可以利用资源指引表明主题的不同属性,并引导用户链接到有关的数字资源或其它形式资源的所在地。前文对黄梅戏主题进行了主题类型的划分,下面选取其中几个主题类型进行属性和资源类型的描述,资源指引说明如表 2 所示:

表 2 资源指引说明

主题类型	属性	资源类型
剧目	正式题名、并列题名、主题词、关键词	文本、图片、音频、视频
音乐风格	风格名称、含义	文本、音频、视频
艺术家	姓名、别名、简介、资源来源、附注	文本、图片、音频、视频、模型
机构	机构名称、地点、简介、规模、年代	文本、图片、模型

在这里,笔者对黄梅戏资源进行了主题类型、关联、资源指引的定义,为后文黄梅戏数字资源主题图的生成和可视化奠定了基础。

4.4 基于主题图的非遗数字资源整合实现

主题图开发工具在主题图开发中具有十分重要的地位,提供主题图的创建、存储、显示和维护等功能。主题图开发工具通常以主题图引擎或 TMAPI (Topic Map API 主题图应用程序接口) 形式存在,OKS、TM4J、tinyTIM、XTM4XMLDB 都是支持 TMAPI 的主题图引擎。笔者主要介绍主题图 OKS 知识开发组件,它也是本文用到的开发工具。OKS (Ontopia Knowledge Suite) 是 Ontopia 公司开发的工具集,它支持建立一个完整的主题图,完全基于 Java 语言,可以为任何应用提供主题图功能,具有稳定性、高效性和先进性^[12]。OKS 体系结构含三层,从下往上分别为存储层、引擎层和应用层。存储层支持基于关系数据库后端 (RDBMS backend) 和基于内存 (In-memory) 的存储技术;引擎层包括主题图引擎、全文本查询引擎;应用层主要包括主题图导航 Navigator、主题图浏览器 Omnigator 和主题图可视化工具 Vizigator。

在构建黄梅戏主题库后,就可以利用 OKS 中的 Ontopoly 进行黄梅戏数字资源主题图的编辑,然后利用 Omnigator 进行浏览,最后通过 Vizigator 实现主题图的可视化。其中,Ontopoly 包含 Description、Export、Instance、Ontology 4 个模块,Description 是对主题图的总描述,即需要对所构建主题图的目的、意义、资源来源等进行说明;Export 可以实现主题图的输出,即可根据需求将所构建的主题图以 XTM、RDF 等不同的格式保存到本地计算机;Ontology 则提供定义主题类型、资源类型、关联类型、角色类型和名称类型的功能;Instance 模块用来定义主题类型的实例。

4.4.1 黄梅戏数字资源主题图的生成与浏览 Ontopoly 提供定义主题类型、资源类型、关联类型、角色类型和名称类型的功能,在 Ontopoly 工具中,依据上文的分析,把黄梅戏数字资源相关信息依次输入。

Omnigator 是 OKS 中用来浏览主题图导航框架的一种免费工具,它以 HTML 的格式输出主题图,在 Omnigator 界面,用户能够直接查看主题之间的关联关系和资源指引。图 3 和图 4 分别展示了用 Omnigator 浏览“Huangmeixi”主题图和传承人之一的“严凤英”主题类型的浏览结果。

4.4.2 黄梅戏数字资源主题图可视化 Navigator 是 OKS 中用来浏览所构建主题图的可视化组件,主题图

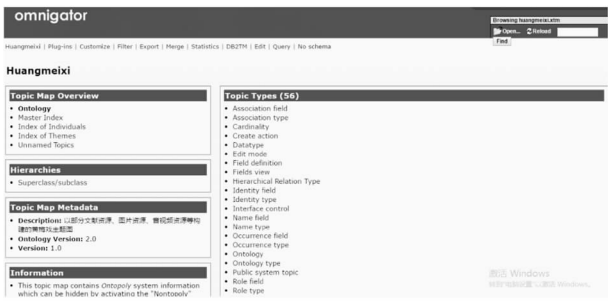


图3 利用 Omnigator 浏览“Huangmeixi”主题图



图4 利用 Omnigator 浏览“严凤英”主题图

可视化展示了利用主题图关联的各主题间的关系形成的一个表达语义的网状结构^[13]。图5和图6分别展现了“严凤英”主题实例和电影“女驸马”主题实例的主题图可视化结果,其中不同的颜色表示不同的关联类型,用户在浏览过程中能选择有兴趣的主题继续追踪,了解更多的相关知识,获得良好的交互体验。

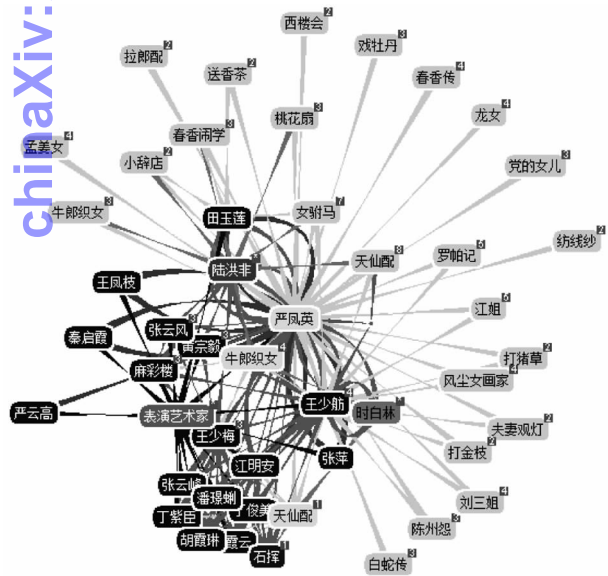


图5 “严凤英”主题实例可视化

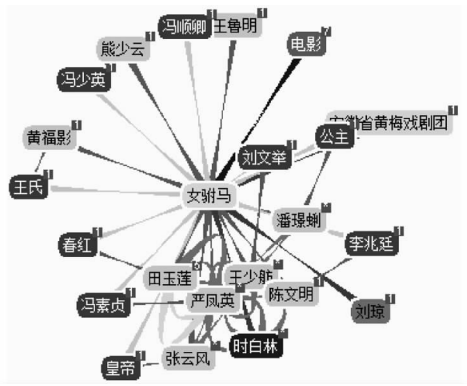


图6 “女驸马”主题实例可视化

英和其他艺术家之间的师徒关系、合作关系等。通过分析可知,构建非遗数字资源主题图将从以下几个方面改善我国目前非遗数字资源的组织情况:

第一,目前我国非遗数字资源还停留在数字化保护和信息整合阶段,没有解决非遗数字资源之间存在的语义关系的问题。笔者构建非遗数字资源主题图为用户提供了基于语义的导航,当用户浏览一个主题时,可以呈现该主题有关的所有主题,并为用户提供它们之间的语义关系,这样用户在浏览某个主题图时可以很方便地找到和该主题相关的其它感兴趣的主题,并了解主题之间的关联关系。

第二,我国非遗资源数量庞大,资源来源十分广泛,用户在查找资源的过程中经常难以找到所需的资源。而在主题图中,由于主题和资源是分开的,用户不需要考虑资源的位置问题,通过主题图的外部资源指引 URI 直接链接到相关类型的文本、图片、音频、视频资源等。

第三,利用主题图提供的可视化插件,用户可以浏览主题图中所有主题的可视化结果,不同主题之间的关系能够得到直观的展示。

5 结语

笔者以非遗数字资源中的黄梅戏为例,利用主题图工具 OKS Sampers 的 Ontopoly 功能构建黄梅戏主题图,构建主题图之后利用 Omnigator 浏览相关主题,并通过 Vizigator 工具实现了黄梅戏主题图的可视化,验证了主题图技术应用于非遗数字资源的可行性。将主题图与非遗数字资源相结合,一方面主题图中的主题和关联允许灵活地定义复杂的语义关系,为非遗数字资源的整合提供一个基于语义的接口,用户可以实现基于语义的非遗数字资源查询检索,同时,可以利用主题图的资源指引功能达成非遗数字资源与主题间的链

接,通过这种映射关系实现对广泛分布的异构非遗数字资源的索引,能有效解决目前非遗数字资源的信息孤岛问题;另一方面主题图将资源域和主题域分离开来,同样的主题可以指向不同的非遗数字资源,不同的主题也可以指向相同的非遗数字资源,这样用户就能查看不同视角的视图,从而满足用户不同的利用需求。但因为非遗数字资源与主题图的结合仍是一个新的研究领域,目前的研究主要处于小数据量研究和理论研究阶段,未来的研究应在大数据量的基础上实现基于主题图的非遗数字资源的信息服务,如非遗数字资源的主题图动态导航、个性化信息检索等功能。

参考文献:

- [1] 王玉桂. 非物质文化遗产的活态特性及其遗产保护[EB/OL]. [2016 - 12 - 07]. <http://www.mzb.com.cn/html/report/141120591-1.htm>.
- [2] 王巨山. 非物质文化遗产的特征及其保护的再认识[J]. 社会科学辑刊,2006(5):165 - 167.
- [3] 王蒙,许鑫. 主题图技术在非物质文化遗产信息资源组织中的应用研究——以京剧、昆曲为例[J]. 图书情报工作,2015,59(14):15 - 21.
- [4] 胡舰. 土家音乐虚拟文化长廊的构建与实现[D]. 武汉:华中师范大学,2016.
- [5] ISO/IEC 13250:2003 Information technology - GML applications -

topic maps[EB/OL]. [2016 - 12 - 17]. <https://www.iso.org/obp/ui/#iso:std:iso-iec:13250:ed-2:v1:en>.

- [6] PEPPER S. The TAO of topic maps[EB/OL]. [2016 - 12 - 22]. <http://www.ontopia.net/topicmaps/materials/tao.html#d0e632>.
- [7] 邓敏. 基于主题图的标签语义挖掘研究[D]. 武汉:华中师范大学,2014:37 - 44.
- [8] 秦铁辉,郭延吉,孙琳. 信息时代的“全球定位系统”——主题地图[J]. 江西图书馆学刊,2005(1):1 - 3,7.
- [9] 张玉涛,夏立新. 基于主题图的电子政务信息资源整合模型研究[J]. 情报杂志,2009(7):161 - 165.
- [10] 黄梅戏[EB/OL]. [2016 - 12 - 30]. <http://www.ihchina.cn/5/10821.html>.
- [11] 《中国戏曲志》委员会. 中国戏曲志. 安徽卷[M]. 北京:中国 ISBN 中心,1993:366 - 367.
- [12] Product white paper the Ontopia Knowledge Suite[EB/OL]. [2016 - 11 - 18]. <http://xml.coverpages.org/Ontopia-OKSv13.pdf>.
- [13] DING Y, JACOB E K, FRIED M, et al. Upper Tag Ontology (UTO) for integrating social tagging data[J]. Journal of the American Society for Information Science & Technology, 2010, 61(3): 505 - 521.

作者贡献说明:

施旻:撰写论文;
熊回香:指导论文;
陆颖颖:修改与润色论文。

An Empirical Analysis on the Digital Resources of Intangible Cultural Heritage Based on Topic Map

Shi Yi¹ Xiong Huixiang² Lu Yingying²

¹ Wuhan University Library, Wuhan 430072

² Central China Normal University, School of Information Management, Wuhan 430079

Abstract: [Purpose/significance] Intangible cultural heritage resources have important social value. If the topic map technology can be applied to the digital resources integration of intangible cultural heritage, it will be beneficial to the digital resources protection and inheritance of intangible cultural heritage. [Method/process] In reaction to the digital resources of intangible cultural heritage, this paper first analyzed the theme, relevance and resource indexing, and then, constructed a topic map and visualized it. [Result/conclusion] The research results show that the construction of topic map of digital cultural heritage digital resources can provide semantic integration, visually present the association of different topics to users, as well as provide visual navigation for users.

Keywords: topic map intangible culture heritage digital recourse visualization